

OPEN

Detecting coalitions by optimally partitioning signed networks of political collaboration

Samin Aref^{1,2*} & Zachary Neal³

We propose new mathematical programming models for optimal partitioning of a signed graph into cohesive groups. To demonstrate the approach's utility, we apply it to identify coalitions in US Congress since 1979 and examine the impact of polarized coalitions on the effectiveness of passing bills. Our models produce a globally optimal solution to the NP-hard problem of minimizing the total number of intra-group negative and inter-group positive edges. We tackle the intensive computations of dense signed networks by providing upper and lower bounds, then solving an optimization model which closes the gap between the two bounds and returns the optimal partitioning of vertices. Our substantive findings suggest that the dominance of an ideologically homogeneous coalition (i.e. partisan polarization) can be a protective factor that enhances legislative effectiveness.

We propose a general method for identifying cohesive groups in signed networks (networks with positive and negative edges), and apply it to political networks, which have become a common focus in complex network analysis¹⁻³. Specifically, we examine signed networks of political collaboration and opposition to identify the members of polarized coalitions in the US Congress, then use these coalitions to examine the impact of polarization on effectiveness in passing bills.

In legislative bodies where most pairs of legislators co-sponsor bills, a network of who co-sponsors with whom becomes a highly dense network which would not be suitable for studying political alliances, coalitions, and polarization. Instead, we use signed networks⁴ created based on significantly *many* and significantly *few* co-sponsorships as two types of edges with opposite nature where a stochastic degree sequence model (SDSM)⁵ is used as the null model, to define thresholds of “many” and “few”. Previous research⁶ on the same data has shown an increase in polarization in the US Congress when measured by the triangle index, which provides a locally-aggregated index of polarization based on structural balance⁷. However, the triangle index only measures the level of balance and polarization, but does not identify the members of the political coalitions that are polarized. For this we turn to the frustration index⁸⁻¹⁰ (also known as the line index of balance¹¹), which optimally partitions a signed graph into two opposing but internally cohesive “coalitions”¹². Substantively, these coalitions¹³ represent groups of legislators who sponsor significantly many bills with each other (i.e. are political allies), but who sponsor significantly few bills with those in the other coalition (i.e. are political enemies). In our analyses of legislative effectiveness, we focus on the level of partisanship within the largest, and therefore controlling, coalition.

Computing the frustration index is an NP-hard problem¹⁴, and so is the equivalent partitioning problem that deals with minimizing the total number of intra-group negative and inter-group positive edges. The optimality of a numerical solution to an instance of an optimization problem depends on the function under optimization. Most studies on this topic use heuristic methods for partitioning signed networks under similar objectives¹⁵⁻¹⁸. These methods are not guaranteed to provide the optimal solution or even its approximation within a constant factor^{14,19}, but can potentially be implemented on larger networks.

Computing the exact value of the frustration index, in principle, involves searching among all possible ways to partition a given signed network into $k \leq 2$ groups in order to find the partitioning which minimizes the total number of intra-group negative and inter-group positive edges. We propose a new method for tackling the intensive computations by providing upper and lower bounds for this number, then solving an optimization model

¹Laboratory of Digital and Computational Demography, Max Planck Institute for Demographic Research, 18057, Rostock, Germany. ²School of Computer Science, University of Auckland, 1142, Auckland, New Zealand.

³Department of Psychology, Michigan State University, East Lansing, MI, 48824, USA. *email: sare618@aucklanduni.ac.nz

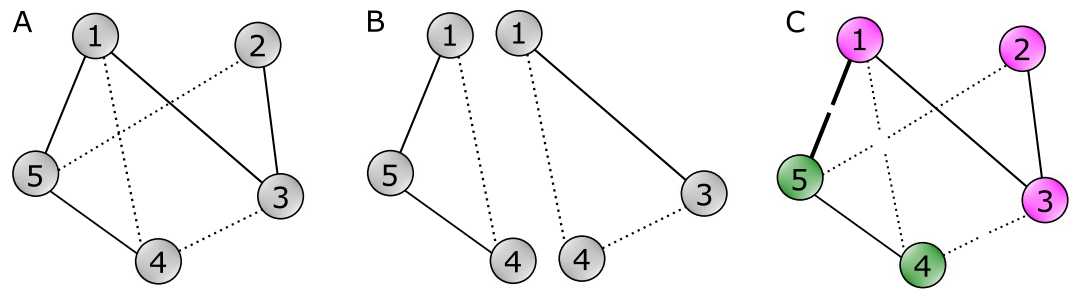


Figure 1. (A) An example signed network. (B) Evaluating balance using triangles. (C) Evaluating balance using frustration.

which closes the gap between the two bounds and returns the exact value of frustration index alongside the optimal partitioning of vertices.

Signed Graph and Balance Theory Preliminaries

In this section, we recall some basic definitions of signed graphs and balance theory.

Signed graphs. We consider an undirected signed graph $G = (V, E, \sigma)$ where V and E are the sets of vertices and edges respectively, and σ is the sign function $\sigma: E \rightarrow \{-1, +1\}$. Graph G contains $|V| = n$ nodes. The set E of edges contains m^- negative edges and m^+ positive edges adding up to a total of $|E| = m = m^+ + m^-$ edges. The *signed adjacency matrix* and the *unsigned adjacency matrix* are denoted by \mathbf{A} and $|\mathbf{A}|$ respectively. Their entries are defined in Eqs. (1) and (2).

$$a_{uv} = \begin{cases} \sigma_{(u,v)} & \text{if } (u, v) \in E \\ 0 & \text{if } (u, v) \notin E \end{cases} \quad (1)$$

$$|a_{uv}| = \begin{cases} 1 & \text{if } (u, v) \in E \\ 0 & \text{if } (u, v) \notin E \end{cases} \quad (2)$$

Balance and cycles. A *cycle* of length k in G is a sequence of nodes $v_0, v_1, \dots, v_{k-1}, v_k$ such that for each $i = 1, 2, \dots, k$ there is an edge from v_{i-1} to v_i and the nodes in the sequence except for $v_0 = v_k$ are distinct. The *sign* of a cycle is the product of the signs of its edges. A cycle with negative (positive) sign is unbalanced (balanced). A balanced network (graph) is one with no negative cycles.

Balance theory is conceptualized by Heider in the context of social psychology²⁰. It was then formulated as a set of graph-theoretic conditions by Cartwright and Harary²¹ which define a signed graph to be balanced if all its cycles are positive. Cartwright and Harary also introduce measuring the level of balance using, among other indices, the fraction of positive cycles [21, page 288]. Three years later, Harary suggested using frustration index¹¹ (under a different name); a measure which satisfies key axiomatic properties¹⁰, but has been underused for decades due to the complexity involving its computation^{19,22,23}.

Evaluating Balance and Frustration

In this section, we explain our computational approach to analyzing signed networks by providing brief definitions and discussions on measuring balance, frustration and partitioning, and graph optimization models.

Measuring partial balance. Signed networks representing real data are often unbalanced, which motivates measuring the intermediate level of partial balance¹⁰. The first measure we use is *triangle index* denoted by $T(G)$ which equals the fraction of positive cycles of length 3^{21,24}. We use Eq. (3) suggested in⁷ for computing triangle index, $T(G)$, in which $\text{Tr}(\mathbf{A})$ denotes the trace (sum of diagonal entries) of \mathbf{A} .

$$T(G) = \frac{\text{Tr}(\mathbf{A}^3) + \text{Tr}(|\mathbf{A}|^3)}{2\text{Tr}(|\mathbf{A}|^3)} \quad (3)$$

The other measure we use is the *normalized frustration index*¹⁰ denoted by $F(G)$ which is based on normalizing the minimum number of edges whose removal results in a balanced graph^{8,11,25}.

Figure 1(A) shows an example signed graph in which the three dotted lines represent negative edges and the four solid lines represent positive edges. The level of balance in this signed graph can be evaluated using triangles (B) or frustration (C). The former approach, (B), involves identifying triangle 1-4-5 as unbalanced and triangle 1-3-4 as balanced leading to the numeric index $T(G) = 1/2$. The latter approach, (C), involves finding a partitioning of vertices $\{\{1, 2, 3\}, \{4, 5\}\}$ (shown by green and purple colors in Fig. 1) which minimizes the total number of intra-group negative and inter-group positive edges to 1 (only edge (1, 5) according to this partitioning). Note that removing edge (1, 5) leads to a balanced signed graph.

Frustration and partitioning. Given signed graph $G = (V, E, \sigma)$, we can partition V into two subsets: X and $V \setminus X$. We let binary variable x_i denote the subset which node i belongs to under partitioning $\{X, V \setminus X\}$, where $x_i = 1$ if $i \in X$ and $x_i = 0$ otherwise.

A positive edge $(i, j) \in E^+$ is said to be frustrated if its endpoints i and j belong to different subsets ($x_i \neq x_j$). A negative edge $(i, j) \in E^-$ is said to be frustrated if its endpoints i and j belong to the same subset ($x_i = x_j$). We define the *frustration count* $f_G(X)$ as the number of frustrated edges of G under partitioning $\{X, V \setminus X\}$:

$$f_G(X) = \sum_{(i,j) \in E} f_{ij}(X)$$

where $f_{ij}(X)$ is the frustration state of edge (i, j) , given by

$$f_{ij}(X) = \begin{cases} 0, & \text{if } x_i = x_j \text{ and } (i, j) \in E^+ \\ 1, & \text{if } x_i = x_j \text{ and } (i, j) \in E^- \\ 0, & \text{if } x_i \neq x_j \text{ and } (i, j) \in E^- \\ 1, & \text{if } x_i \neq x_j \text{ and } (i, j) \in E^+. \end{cases} \quad (4)$$

The frustration index of a graph G can be computed exactly by finding partitioning $X^*, V \setminus X^* \subseteq V$ of G that minimizes the frustration count $f_G(X)$, i.e. solving Eq. (5)^{19,22}.

$$L(G) = \min_{X \subseteq V} f_G(X) \quad (5)$$

The normalized frustration index, $F(G)$, is computed based on $L(G)$ and according to Eq. (6) which allows measuring the level of partial balance based on numerical values within the unit interval (m denotes the number of edges).

$$F(G) = 1 - 2L(G)/m \quad (6)$$

One may notice some similarities between the problem of finding communities in unsigned networks^{26–31} and that of partitioning signed networks to minimize the frustration count. One key difference is that in the latter problem for every pair of vertices there are three cases (as opposed to two): a positive edge, a negative edge, or no edge between the two vertices. Due to the differences between objectives of these two problems (minimizing frustration count as opposed to maximizing modularity or other quantities), the partitioning obtained from running community detection algorithms on positive edges of a signed graph will not generally minimize the frustration count.

Recent studies on frustration index and signed networks suggest^{19,22} and implement²³ efficient graph optimization models to compute the frustration index of relatively large (up to 10^5 edges) sparse networks. However, the signed networks we analyze have substantially higher densities compared to the instances in^{19,22,23}. This requires developing new computational models for tackling the intensive computations involved in obtaining the frustration index of dense graphs.

Bounding the frustration index. In this subsection, we discuss obtaining lower and upper bounds for the frustration index. Using these bounds is a way of substantially reducing the running time, but theoretically they are not required.

The linear programming relaxation (LP relaxation) of the binary optimization models in^{19,22} can be used to compute a lower bound for the frustration index. The linear programming model in Eq. (7) is developed for this purpose.

$$\begin{aligned} \min_{x_p, x_{ij}} Y = & \sum_{(i,j) \in E^+} x_i + x_j - 2x_{ij} + \sum_{(i,j) \in E^-} 1 - (x_i + x_j - 2x_{ij}) \\ \text{s. t.} & x_{ij} \leq (x_i + x_j)/2 \quad \forall (i, j) \in E^+ \\ & x_{ij} \geq x_i + x_j - 1 \quad \forall (i, j) \in E^- \\ & x_i + x_{jk} \geq x_{ij} + x_{ik} \quad \forall (i, j, k) \in T \\ & x_j + x_{ik} \geq x_{ij} + x_{jk} \quad \forall (i, j, k) \in T \\ & x_k + x_{ij} \geq x_{ik} + x_{jk} \quad \forall (i, j, k) \in T \\ & 1 + x_{ij} + x_{ik} + x_{jk} \geq x_i + x_j + x_k \quad \forall (i, j, k) \in T \\ & x_i \in [0, 1] \quad \forall i \in V \\ & x_{ij} \in [0, 1] \quad \forall (i, j) \in E \end{aligned} \quad (7)$$

In Eq. (7), $T = \{(i, j, k) \in V^3 | (i, j), (i, k), (j, k) \in E\}$ is the set which contains ordered 3-tuples of nodes whose edges form a triangle in G . The continuous linear programming model in Eq. (7) is developed by combining the LP relaxation of the 0/1 linear model in²² [Subsection 4.3] and the triangle constraints in²² [Subsection 4.4]. It follows from the LP relaxation that the optimal solution Y^* to the model in Eq. (7) is a lower bound for the frustration index $Y^* \leq L(G)$.

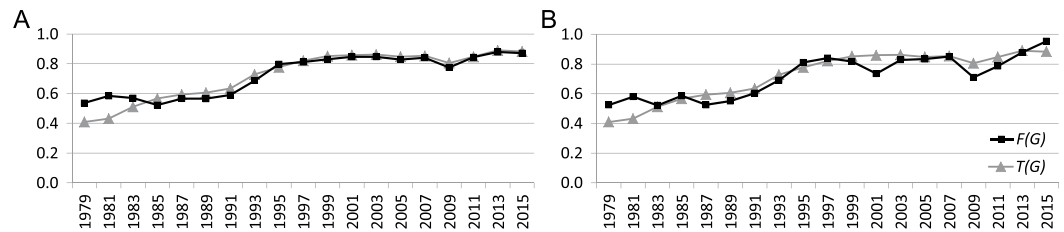


Figure 2. Two measures of partial balance indicating an overall increase in political polarization in (A) US House of Representatives and (B) US Senate over the time period 1979–2016.

Any given partitioning $\{X, V \setminus X\}$ for signed graph G is associated with a frustration count $f_G(X)$ which is by definition (as in Eq. (5)) an upper bound for the frustration index

$$f_G(X^*) = L(G) \leq f_G(X) \quad \forall X \subseteq V.$$

We use a specific partitioning $\{X', V \setminus X'\}$ as a starting point to “warm-start” the algorithm for computing the frustration index. Partitioning $\{X', V \setminus X'\}$ groups nodes into two subsets based on the party affiliation of legislators. To be more precise, for node i which represents a legislator, decision variable x_i is given initial value 0 if the reciprocal legislator is a Democrat and x_i is given initial value 1 otherwise.

Computing the frustration index. After bounding the frustration index, we use the binary linear programming model in Eq. (8) which minimizes the number of frustrated edges.

$$\begin{aligned} \min_{x_i, f_{ij}} Z &= \sum_{(i,j) \in E} f_{ij} \\ \text{s. t. } f_{ij} &\geq x_i - x_j \quad \forall (i, j) \in E^+ \\ f_{ij} &\geq x_j - x_i \quad \forall (i, j) \in E^+ \\ f_{ij} &\geq x_i + x_j - 1 \quad \forall (i, j) \in E^- \\ f_{ij} &\geq 1 - x_i - x_j \quad \forall (i, j) \in E^- \\ \sum_{(i,j) \in E} f_{ij} &\geq Y^* \\ x_i &\in \{0, 1\} \quad \forall i \in V \\ f_{ij} &\in \{0, 1\} \quad \forall (i, j) \in E \end{aligned} \quad (8)$$

The binary variables of the model are $f_{ij} \quad \forall (i, j) \in E$ which denotes frustration of edge (i, j) and $x_i \quad \forall i \in V$ which denotes the subset of node i . To warm-start the algorithm which solves Eq. (8), we initialize x_i variables based on partitioning $\{X', V \setminus X'\}$. The model in Eq. (8) is developed by combining the XOR model in¹⁹ [Subsection 3.2] with an additional constraint to incorporate the lower bound Y^* obtained from Eq. (7). We implement the speed-up techniques discussed in¹⁹ and solve the binary linear programming model in Eq. (8) using *Gurobi* solver (version 8.0)³² on a virtual machine with 32 Intel Xeon CPU E5-2698 v3 @ 2.30 GHz processors and 32 GB of RAM running 64-bit Microsoft Windows Server 2012 R2 Standard.

Results

In this section, we provide the results of analyzing balance and frustration in signed networks of US Congress legislators.

Partial balance, frustration, and optimal partitioning. We evaluate the level of partial balance using two different methods. Figure 2 illustrates partial balance in the signed networks of the US Congress over time measured by the triangle index and normalized frustration index. Values of the two measures, the triangle index $T(G)$ and the normalized frustration index $F(G)$, are highly correlated (correlation coefficients are 0.95 and 0.91 respectively for House and Senate networks) and both show relatively high levels of partial balance which have increased in the time period 1979–2016. The results in Fig. 2 indicate an increase in the polarization of both chambers of US Congress, which is in accordance with the literature^{6,33–36}. Although the triangle index $T(G)$ and the normalized frustration index $F(G)$ capture very similar information concerning the level of partial balance, only the computation of $F(G)$ also provides the partitioning that minimizes the sum of intra-group negative and inter-group positive edges.

Solving the continuous optimization model in Eq. (7) and the discrete (binary) optimization model in Eq. (8) requires intensive computations for large instances such as signed networks of the House. Given the size and density of these instances, the models in Eqs. (7) and (8) have thousands of variables and possibly millions of constraints requiring a high performance computer taking advantage of parallel computing capabilities²³.

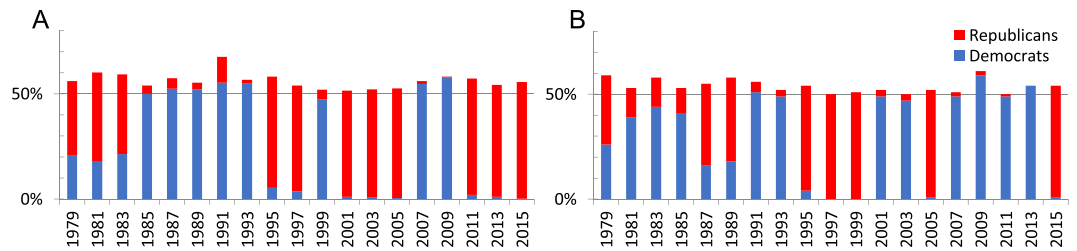


Figure 3. The number of legislators from the two main parties who belong to the controlling coalition indicating an increase in the partisan homogeneity of the controlling coalition in (A) US House of Representatives and (B) US Senate over the time period 1979–2016.

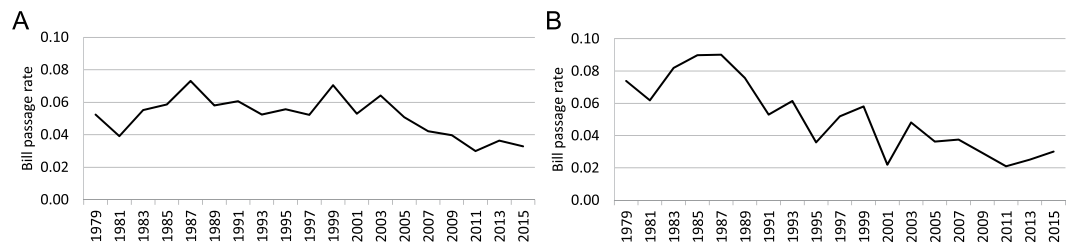


Figure 4. The fraction of bills that eventually become law (bill passage rate) indicating legislative effectiveness in (A) US House of Representatives and (B) US Senate over the time period 1979–2016.

For example, the signed graph instance of the 113th House session has $m = 75,771$ edges and $|T| = 7,102,625$ triangles which result in a total of $m + 4|T| = 28,486,271$ constraints for the model in Eq. (7). Gurobi solver takes 5300 seconds (around 1.5 hours) to solve the model in Eq. (7) to global optimality and return Y^* , the lower bound for the frustration index. For the same instance, the discrete optimization model in Eq. (8) has $n + m = 76,218$ binary variables and $2m + 1 = 151,543$ constraints. This large instance takes 43,523 seconds (around 12 hours) for Gurobi to reach global optimality and return the frustration index and the partitioning of the nodes. In total for the 113th session of the House, it takes 48,823 seconds (around 13.5 hours) to compute the exact value of the frustration index which is the longest solve time among all instances. The average computation time for House instances is 17,763 seconds (around 5 hours) and the standard deviation is 15,411 seconds (around 4.5 hours). For Senate instances, the average computation time is 4 seconds and the standard deviation is 6 seconds.

Using the optimal values of the x_i variables obtained by solving the discrete optimization model in Eq. (8), we partition nodes of each network into two groups (subsets X^* , $V \setminus X^*$). For each signed network, either X^* or $V \setminus X^*$ has the larger set cardinality and therefore represents the largest coalition for the corresponding session.

We evaluate the composition of the largest and therefore controlling coalitions in each session and chamber based on the party affiliation of its legislators. Figure 3 illustrates the number of legislators from the two main political parties in the controlling coalitions of the US Congress. As it can be seen in Fig. 3, the controlling coalitions have become more homogeneous (i.e. partisan) over the time period 1979–2016.

Legislative effectiveness and polarization in the US congress. Within the field of comparative US politics, two topics attract particular attention at the federal level: legislative effectiveness and political polarization. Legislative effectiveness refers to the ability of individual legislators^{37,38}, or of an entire legislative body³⁹, to advance their agenda, typically by facilitating the passage of legislation. Political polarization (when applied to elected officials or “elites”) refers to the formation of non-overlapping ideologically homogeneous groups^{6,33}. When these groups mirror political party affiliations, it is also called partisan polarization. For several decades, legislative effectiveness in the US has declined (as illustrated in Fig. 4), while partisan polarization has increased⁶. These trends have led many to hypothesize that they are related, and specifically that “unified party control has [not] been legislatively more productive than divided party control” [40, xii]. Based on the legislative process used by the US Congress, it might be expected that a chamber’s bills are more likely to become law when the controlling party holds a larger majority, because its members can form a voting bloc. However, the analysis in the next section suggests that that changes in bill passage rates are better explained by the partisanship of a chamber’s largest coalition.

Mediation in bill passage. Using a bivariate linear regression, we find that the percentage of bills introduced in a chamber that become law (passage rate) significantly declines over time. The passage rate has declined in the House by an average of 0.11 percentage points each session ($\beta = -0.528$, $p < 0.05$), and in the Senate by an average of 0.35 percentage points each session ($\beta = -0.852$, $p < 0.01$; see Fig. 5(A)). We report standardized coefficients (β) to facilitate cross-model comparisons. The percentage point changes are unstandardized bivariate regression coefficients, reported here for context. These variations in passage rate are not simply a function of the

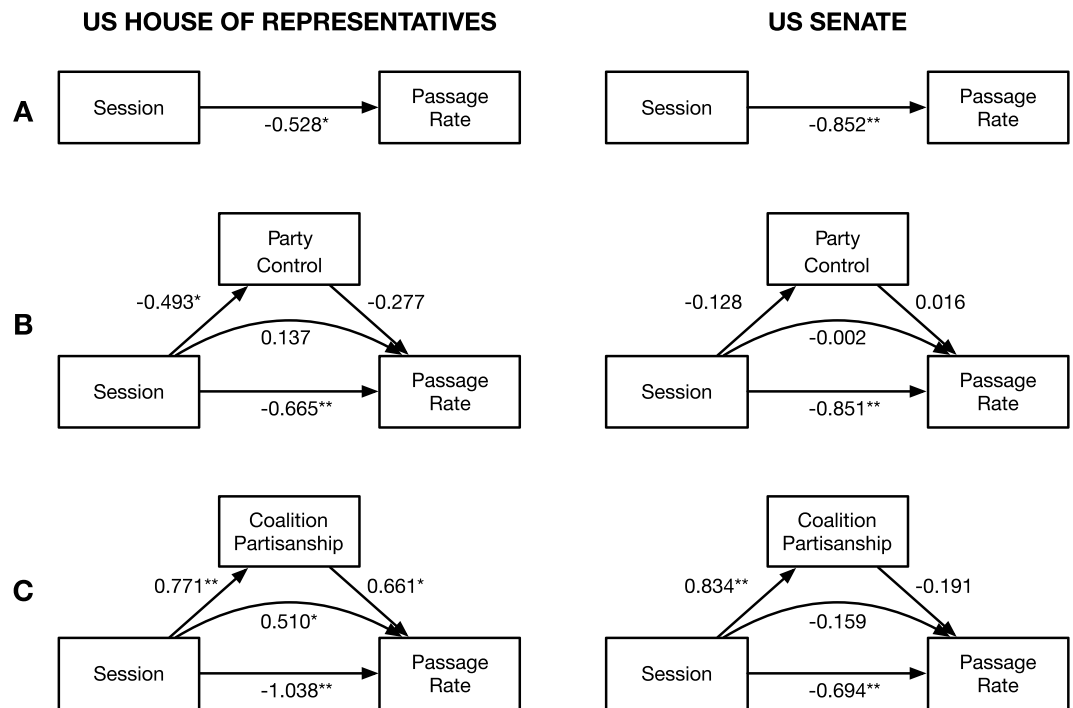


Figure 5. Predicting the rate of bill passage in the US House of Representatives and Senate; Standardized coefficients are reported; ** $p < 0.01$, * $p < 0.05$. (A) Over time, the passage rate has declined. (B) The decline is not mediated by changes in the size of a party's majority in a chamber. (C) In the House, but not the Senate, it is mediated by the partisan homogeneity in the controlling coalition.

total number of bills introduced (House: $r = -0.29$, $p = 0.22$; Senate: $r = -0.08$, $p = 0.74$), which exhibit no association with time (House: $r = -0.34$, $p = 0.15$; Senate: $r = 0.19$, $p = 0.43$; see Tables S2 and S4).

To investigate possible explanations for variations in passage rate, we estimate two separate structural equation models for each chamber. A commonsense model tests the expectation that when the majority party holds a larger numerical majority, they should have greater success passing bills⁴¹. The key variable in this model, *party control*, is defined as the absolute difference between the number of Republicans and Democrats. Computing party control does not require any information about the legislators' network. We find no support for this model; party control does not mediate the relationship between time and passage rate (see Fig. 5(B)).

A more nuanced model tests the expectation that when the controlling coalition is more partisan and thus more ideologically unified, it will have greater success passing bills. The key variable in this model, *coalition partisanship*, is defined as the fraction of non-independent members in the largest coalition that affiliate with that coalition's dominant political party (see Fig. 3). We compute coalition partisanship by applying the partition method described above to a signed network of political collaboration and opposition. We find support for this model in the US House, but not the US Senate (see Fig. 5(C)). Specifically, we find that in the House, the partisan homogeneity of the controlling coalition has increased over time ($\beta = 0.771$, $p < 0.01$), which is consistent with past findings concerning polarization⁶, and that coalition homogeneity increases passage rates ($\beta = 0.661$, $p < 0.05$), which is consistent with our expectation about the impact of ideological unity. Together, these effects imply a significant and positive indirect effect of time on the passage rate ($\beta = 0.510$, $p < 0.05$), mediated by coalition partisanship. Thus, the observed decline in bill passage rates in the US House would have been worse (direct effect: $\beta = -1.038$, $p < 0.01$), but was mitigated by increasingly ideologically homogeneous coalitions, which are a protective factor against declines in legislative effectiveness.

Summary and Conclusions

In this study we proposed a general method for identifying internally cohesive opposing coalitions in signed networks of legislators based on structural balance theory, then applied this method to identify opposing coalitions in the US Congress, showing that these coalitions' partisanship can explain changes in legislative effectiveness better than political parties. Based on this analysis, we offer a series of substantive and methodological conclusions.

Consistent with prior studies^{6,33–36}, we find that polarization has increased in both the US Senate and US House of Representatives, and that this polarization has largely mirrored partisan divisions along political party lines. We operationalized polarization using the level of a signed graph's structural balance, and therefore measure what⁶ calls "strong polarization," but have used two different measures of balance. We find that the two measures are highly correlated and both support the conclusion of increasing polarization.

The triangle index is easy to compute, but provides only a locally-aggregated measure of a graph's level of balance. In contrast, computing the frustration index is difficult, but it provides not only a global measure of a graph's

level of balance, but also the optimal partitioning of vertices into internally cohesive but mutually antagonist groups. We have demonstrated a practical method for computing the exact value of frustration index and identifying the optimal partition in dense graphs of $|E| \gg 50000$ that involves first obtaining upper and lower bounds, using exogenous node properties (e.g. legislators' political party affiliations), and solving a large-scale binary linear programming model. In the context of legislative networks, this method allows us to identify the most cohesive coalitions of legislators under conditions of balance theory.

Although our computational innovations make the identification of internally cohesive opposing coalitions practically feasible, we must also demonstrate that these coalitions are more informative than other simpler grouping possibilities. In the legislative context, we show that the partisan composition of these cohesive coalitions better explains the declining legislative effectiveness in the US House of Representatives than simply examining legislators' political party affiliations. This affirms Mayhew's claim that "no theoretical treatment of the United States Congress that posits parties as analytic units will go very far" [42, p.27] but goes a step further by identifying an alternative analytic unit – internally cohesive opposing coalitions – that does have explanatory power. Importantly, coalitions appear useful only for explaining the legislative effectiveness of the House, but not the Senate. However, this is also consistent with existing political science theory that "the lack of majority control of [procedural] processes in the Senate negates the possibility of significant party [or other group-based] effects in that body" [43, p.7]. Therefore, in general terms, our empirical findings suggest that in legislative bodies where a sufficiently large group of legislators can influence procedural processes, the composition of the largest coalition is more important than the size of the majority party's majority. This is perhaps obvious in parliamentary systems where multi-party coalition forming is essential, but is noteworthy in the non-parliamentary US Congress.

These conclusions have some significant implications for both the future study of signed networks, and of the link between polarization and legislative effectiveness. First, by providing a practical method for computing the frustration index of relatively dense graphs, we hope to move the study of signed graphs beyond merely determining the level of balance, and toward the study of how the composition of mostly opposing groups impact other network dynamics. Second, our empirical findings suggest that research on polarization and its impact on the legislative process should look beyond political parties and partisanship to more subtle but influential forms of coordination, such as internally cohesive coalitions which are antagonist towards one another.

Materials and Methods

Relations of collaboration and opposition between elected officials are difficult to collect directly because politicians have limited time to participate in surveys and have good reasons to conceal their true political relations. Therefore, studies of elected officials' political networks typically measure these relations indirectly, using bipartite projections focusing on their co-sponsorship of bills⁴⁴, co-voting on bills^{36,45,46}, co-membership on committees⁴⁷, and co-attendance at press events⁴⁸. For a range of substantive reasons noted by⁶ (e.g. relatively few bills are actually voted on, committee memberships are driven by such non-ideological factor such as seniority), we examine political relations from bill co-sponsorship.

Specifically, we use a signed network of inferred political relations among the members of the US House of Representatives, and among the members of the US Senate, in each session of Congress from 1979 to 2016 (96th session – 114th session). The process for creating these signed networks is described in detail by⁶ and they are available in a public *Figshare* data repository⁴.

Inferring signed networks from co-sponsorship data. Importantly, all pairs of legislators co-sponsor at least some of the same bills, so we know that the mere existence of some co-sponsorships does not imply they collaborate, and that some number of co-sponsorships can actually indicate avoidance. In previous work⁶, a stochastic degree sequence model (SDSM)⁵ is used to define thresholds of significantly few and significantly many co-sponsorships by building the empirical sampling distribution of two legislators' joint co-sponsorships under a null model in which each legislator co-sponsored approximately the same number of bills and each bill received approximately the same number of co-sponsorships (i.e. holding approximately constant the legislator and bill degree sequence). To be more specific, given a bipartite graph B , Monte Carlo methods can be used to generate probability distributions BB'_{ij} when $Pr(B_{ij} = 1)$ is a function of the row and column marginals of B ⁶. Decisions about whether a given dyad represents significantly few or significantly many co-sponsorships are made by comparing their observed number of joint co-sponsorships to the empirical sampling distribution using a two-tailed $\alpha = 0.05$ threshold. For example, Fig. 6 shows that Rep. Earl Blumenhauer (D-OR3) and Rep. Sheila Jackson-Lee (D-TX18) were observed to have co-sponsored 242 of the same bills (dashed vertical line). The magnitude of joint co-sponsorships, and the fact that both representatives are Democrats, might lead one to conclude that they are collaborating. However, the shaded distribution shows the expected number of joint co-sponsorships under the SDSM null model in which each representative randomly chooses which bills to co-sponsor. Comparing these representatives' observed number of joint co-sponsorships to the null model expectation, we find that they co-sponsor significantly fewer of the same bills than would be expected at random and therefore define the edge between them as negative.

This approach differs from other methods of reducing weighted graphs to binary or signed graphs^{49,50} because it explicitly incorporates information from the original bipartite data (i.e. legislators linked to bills), thereby ensuring it is not lost when these data are projected as a unipartite graph. Additionally,⁶ extracts signed backbone networks rather than the weighted bipartite projections because the weights in those projections are distorted by heterogeneity in the bipartite degree sequences (i.e. some legislators sponsor many bills, others sponsor few^{5,51}).

Although data on earlier sessions are available, they were excluded because prior to the 96th session, House rules imposed a limit of 25 co-sponsors per bill, which artificially distorts co-sponsorship patterns and limits the usefulness of these data for inferring political networks⁵². These data do not distinguish between a bill's "sponsor"

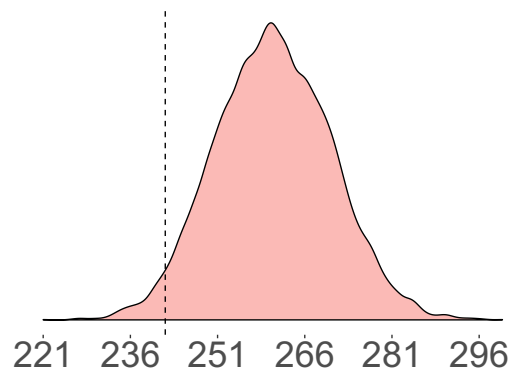


Figure 6. The observed number of co-sponsorships and the null model sampling distribution obtained using the SDSM for Rep. Earl Blumenhauer (D-OR3) and Rep. Sheila Jackson-Lee (D-TX18).

and its “co-sponsors” because the former is simply the legislator whose name appears first in a potentially long list of legislators responsible for the bill’s introduction.

Although these data represent a time-series of legislative interactions, we examine the networks cross-sectionally for two reasons. First, there are a large number of joiners and leavers in each new session as incumbents lose their seats, freshmen join Congress, or representatives become senators, making most dynamic models impractical to estimate. Second, although some political relationships develop over long periods of time, the effectiveness of any particular session of Congress can be evaluated independently.

Data availability

All the data and codes used in this study are publicly available with links and descriptions provided in the Supplementary Information.

Received: 13 November 2019; Accepted: 6 January 2020;

Published online: 30 January 2020

References

- Ribeiro, H. V., Alves, L. G. A., Martins, A. F., Lenzi, E. K. & Perc, M. The dynamical structure of political corruption networks. *Journal of Complex Networks* **6**, 989–1003, <https://doi.org/10.1093/comnet/cny002> (2018).
- Colliri, T. & Zhao, L. Analyzing the bills-voting dynamics and predicting corruption-convictions among Brazilian congressmen through temporal networks. *Scientific Reports* **9**, 16754, <https://doi.org/10.1038/s41598-019-53252-9> (2019).
- Faustino, J., Barbosa, H., Ribeiro, E. & Menezes, R. A data-driven network approach for characterization of political parties’ ideology dynamics. *Applied Network Science* **4**, 48, <https://doi.org/10.1007/s41109-019-0161-0> (2019).
- Neal, Z. A Sign of the Times: Dataset of US Congress signed network backbones from co-sponsorship data, 1973–2016. figshare <https://doi.org/10.6084/m9.figshare.8096429> (2019).
- Neal, Z. The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Social Networks* **39**, 84–97, <https://doi.org/10.1016/j.socnet.2014.06.001> (2014).
- Neal, Z. A sign of the times? Weak and strong polarization in the U.S. Congress, 1973–2016. *Social Networks* **60**, 103–112, <https://doi.org/10.1016/j.socnet.2018.07.007> (2020).
- Terzi, E. & Winkler, M. A spectral algorithm for computing social balance. In Frieze, A., Horn, P. & Pralat, P. (eds) *Proceedings of International Workshop on Algorithms and Models for the Web-Graph*, WAW 2011, 1–13, https://doi.org/10.1007/978-3-642-21286-4_1 (Springer, 2011).
- Zaslavsky, T. Balanced decompositions of a signed graph. *Journal of Combinatorial Theory, Series B* **43**, 1–13, [https://doi.org/10.1016/0095-8956\(87\)90026-8](https://doi.org/10.1016/0095-8956(87)90026-8) (1987).
- Facchetti, G., Iacono, G. & Altafini, C. Computing global structural balance in large-scale signed social networks. *Proceedings of the National Academy of Sciences* **108**, 20953–20958, <https://doi.org/10.1073/pnas.1109521108> (2011).
- Aref, S. & Wilson, M. C. Measuring partial balance in signed networks. *Journal of Complex Networks* **6**, 566–595, <https://doi.org/10.1093/comnet/cnx044> (2018).
- Harary, F. On the measurement of structural balance. *Behavioral Science* **4**, 316–323, <https://doi.org/10.1002/bs.3830040405> (1959).
- Harary, F. & Kabell, J. A. A simple algorithm to detect balance in signed graphs. *Mathematical Social Sciences* **1**, 131–136, [https://doi.org/10.1016/0165-4896\(80\)90010-4](https://doi.org/10.1016/0165-4896(80)90010-4) (1980).
- Riker, W. H. *The theory of political coalitions* (Yale University Press, 1962).
- Hüffner, F., Betzler, N. & Niedermeier, R. Separator-based data reduction for signed graph balancing. *Journal of Combinatorial Optimization* **20**, 335–360, <https://doi.org/10.1007/s10878-009-9212-2> (2010).
- Gong, M., Cai, Q., Ma, L., Wang, S. & Lei, Y. Network structure balance analytics with evolutionary optimization. In *Computational Intelligence for Network Structure Analytics*, 135–199, https://doi.org/10.1007/978-981-10-4558-5_4 (Springer, 2017).
- Traag, V. A., Doreian, P. & Mrvar, A. Partitioning signed networks. In Doreian, P., Batagelj, V. & Ferligoj, A. (eds) *Advances in network clustering and blockmodeling*, chap. 8 (Wiley-Interscience, 2018).
- Hua, J., Yu, J. & Yang, M.-S. Fast clustering for signed graphs based on random walk gap. *Social Networks*, <https://doi.org/10.1016/j.socnet.2018.08.008> (2018).
- Brusco, M. J. & Doreian, P. Partitioning signed networks using relocation heuristics, tabu search, and variable neighborhood search. *Social Networks* **56**, 70–80, <https://doi.org/10.1016/j.socnet.2018.08.007> (2019).
- Aref, S., Mason, A. J. & Wilson, M. C. A modeling and computational study of the frustration index in signed networks. *Networks* **75**, 95–110, <https://doi.org/10.1002/net.21907> (2020).
- Heider, F. Social perception and phenomenal causality. *Psychological Review* **51**, 358–378, <https://doi.org/10.1037/h0055425> (1944).

21. Cartwright, D. & Harary, F. Structural balance: a generalization of Heider's theory. *Psychological Review* **63**, 277–293, <https://doi.org/10.1037/h0046049> (1956).
22. Aref, S., Mason, A. J. & Wilson, M. C. Computing the line index of balance using integer programming optimisation. In Goldengorin, B. (ed.) *Optimization Problems in Graph Theory*, 65–84, https://doi.org/10.1007/978-3-319-94830-0_3 (Springer, 2018).
23. Aref, S. & Wilson, M. C. Balance and frustration in signed networks. *Journal of Complex Networks* **7**, 163–189, <https://doi.org/10.1093/comnet/cny015> (2019).
24. Harary, F. Graphing conflict in international relations. *The papers of the Peace Science Society* **27**, 1–10 (1977).
25. Abelson, R. P. & Rosenberg, M. J. Symbolic psycho-logic: A model of attitudinal cognition. *Behavioral Science* **3**, 1–13, <https://doi.org/10.1002/bs.3830030102> (1958).
26. Kernighan, B. W. & Lin, S. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal* **49**, 291–307, <https://doi.org/10.1002/j.1538-7305.1970.tb01770.x> (1970).
27. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **99**, 7821–7826, <https://doi.org/10.1073/pnas.122653799> (2002).
28. Clauset, A., Newman, M. E. J. & Moore, C. Finding community structure in very large networks. *Physical Review E* **70**, 066111, <https://doi.org/10.1103/PhysRevE.70.066111> (2004).
29. Raghavan, U. N., Albert, R. & Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E* **76**, 036106, <https://doi.org/10.1103/PhysRevE.76.036106> (2007).
30. Cordasco, G. & Gargano, L. Community detection via semi-synchronous label propagation algorithms. In *2010 IEEE International Workshop on: Business Applications of Social Network Analysis (BASNA)*, 1–8, <https://doi.org/10.1109/BASNA.2010.5730298> (2010).
31. Parés, F. et al. Fluid communities: A competitive, scalable and diverse community detection algorithm. In Cherifi, C., Cherifi, H., Karsai, M. & Musolesi, M. (eds) *Complex Networks & Their Applications VI*, 229–240, https://doi.org/10.1007/978-3-319-72150-7_19 (Springer International Publishing, Cham, 2018).
32. Gurobi Optimization Inc. Gurobi optimizer reference manual (2018), www.gurobi.com/documentation/8.0/refman/index.html (date accessed 1 June 2018).
33. Layman, G. C., Carsey, T. M. & Horowitz, J. M. Party polarization in american politics: Characteristics, causes, and consequences. *Annual Review of Political Science* **9**, 83–110, <https://doi.org/10.1146/annurev.polisci.9.070204.105138> (2006).
34. Zhang, Y. et al. Community structure in congressional cosponsorship networks. *Physica A* **387**, 1705–1712, <https://doi.org/10.1016/j.physa.2007.11.004> (2008).
35. Waugh, A. S., Pei, L., Fowler, J. H., Mucha, P. J. & Porter, M. A. Party polarization in congress: A network science approach. *arXiv* 0907.3509 (2011).
36. Moody, J. & Mucha, P. J. Portrait of political party polarization. *Network Science* **1**, 119–121, <https://doi.org/10.1017/nws.2012.3> (2013).
37. Olson, D. M. & Nonidez, C. T. Measures of legislative performance in the U.S. House of Representatives. *Midwest Journal of Political Science* **16**, 269–277, <https://doi.org/10.2307/2110060> (1972).
38. Frantzich, S. Who makes our laws? The legislative effectiveness of members of the U.S. congress. *Legislative Studies Quarterly* **4**, 409–428, <https://doi.org/10.2307/439582> (1979).
39. Volden, C. & Wiseman, A. E. *Legislative effectiveness in the United States Congress: The lawmakers* (Cambridge university press, 2014).
40. Mayhew, D. R. *Divided we govern: Party control, lawmaking, and investigations, 1946–2002* (Yale university press, 2005).
41. Moore, D. W. Legislative effectiveness and majority party size: A test in the indiana house. *The Journal of Politics* **31**, 1063–1079, <https://doi.org/10.2307/2128358> (1969).
42. Mayhew, D. R. *Congress: The Electoral Connection* (Yale university press, 1974).
43. Monroe, N. W., Roberts, J. M. & Rohde, D. W. *Why Not Parties? Party Effects in the United States Senate* (University of Chicago Press, 2008).
44. Fowler, J. H. Legislative cosponsorship networks in the US House and Senate. *Social Networks* **28**, 454–465, <https://doi.org/10.1016/j.socnet.2005.11.003> (2006).
45. Andris, C. et al. The rise of partisanship and super-cooperators in the U.S. House of Representatives. *PloS one* **10**, 1–14, <https://doi.org/10.1371/journal.pone.0123507> (2015).
46. Arinik, N., Figueiredo, R. & Labatut, V. Analysis of roll-calls in the European parliament by multiple partitioning of multiplex signed networks. *Social Networks (in press)* (2019), <https://doi.org/10.1016/j.socnet.2019.02.001> (2018).
47. Porter, M. A., Mucha, P. J., Newman, M. E. J. & Warmbrand, C. M. A network analysis of committees in the U.S. House of Representatives. *Proceedings of the National Academy of Sciences* **102**, 7057–7062, <https://doi.org/10.1073/pnas.0500191102> (2005).
48. Desmarais, B. A., Moscardelli, V. G., Schaffner, B. F. & Kowal, M. S. Measuring legislative collaboration: The Senate press events network. *Social Networks* **40**, 43–54, <https://doi.org/10.1016/j.socnet.2014.07.006> (2015).
49. Serrano, M. Á., Boguñá, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences* **106**, 6483–6488, <https://doi.org/10.1073/pnas.0808904106> (2009).
50. Dianati, N. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E* **93**, 012304, <https://doi.org/10.1103/PhysRevE.93.012304> (2016).
51. Latapy, M., Magnien, C. & Vecchio, N. D. Basic notions for the analysis of large two-mode networks. *Social Networks* **30**, 31–48, <https://doi.org/10.1016/j.socnet.2007.04.006> (2008).
52. Thomas, S. & Grofman, B. The effects of congressional rules about bill cosponsorship on duplicate bills: changing incentives for credit claiming. *Public Choice* **75**, 93–98, <https://doi.org/10.1007/BF01053883> (1993).

Acknowledgements

The authors acknowledge Center for eResearch at University of Auckland for providing access to high-performance computers. There is no funding to be reported for this study.

Author contributions

S.A. developed and implemented new methods and algorithms, obtained the results, and prepared Supplementary information; Z.N. prepared network data, ran statistical models, and analyzed the results; both authors contributed to designing and conducting the research and writing the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-58471-z>.

Correspondence and requests for materials should be addressed to S.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020